Testing Predictive Biases at the Intersection of Race-Ethnicity and Sex: A Multi-Site Evaluation of a Pretrial Risk Assessment Tool

Matthew DeMichele Ph.D.
Primary Affiliation:
Center for Courts and Corrections Research, RTI International, Research Triangle Park, NC, 27709

Ian A. Silver Ph.D.
Primary Affiliation:
Center for Courts and Corrections Research, RTI International, Research Triangle Park, NC, 27709

Ryan M. Labrecque, Ph.D.
Primary Affiliation:
Center for Courts and Corrections Research, RTI International, Research Triangle Park, NC, 27709

Debbie Dawes
Primary Affiliation:
Center for Courts and Corrections Research, RTI International, Research Triangle Park, NC, 27709

Pamela K. Lattimore
Primary Affiliation:
Center for Courts and Corrections Research, RTI International, Research Triangle Park, NC, 27709

Stephen Tueller
Primary Affiliation:
Center for Courts and Corrections Research, RTI International, Research Triangle Park, NC, 27709

Funding: This project is funded by John D. and Laura Arnold Ventures. All of the conclusions derived from the current study are of the authors' alone.

***Authors' Contributions*:** Matthew DeMichele was responsible for developing the research idea and drafting the manuscript. Ian A. Silver was responsible conducting data analysis and drafting the manuscript. Ryan M. Labrecque was responsible drafting the manuscript. Debbie Dawes, Pamela K. Lattimore, and Stephen Tueller were responsible for data collection, data processing, and reviewing the manuscript.

Abstract

Pretrial assessments provide courtroom actors with objective information about the likelihood that someone who has been charged with a criminal offense would engage in criminal behavior if released prior to trial. Although prior research supports the ability of assessments to predict pretrial outcomes, there are concerns that assessments may inadvertently exacerbate racial-ethnic and sex disparities found in the larger criminal legal system. In the current study, we conduct a multi-site predictive bias tests of a widely used pretrial instrument—the *Public Safety Assessment (PSA)*—across six racial-ethnic and sex groups (i.e., White males, White females, Black males, Black females, Hispanic males, Hispanic females). Study results support the PSA as a valid and consistent predictor of failure to appear, new criminal activity, and new violent criminal activity across these six racial-ethnic and sex groups. These findings support the use the PSA and identify several areas for future research.

Keywords: jail, gender, pretrial, PSA, race, pretrial assessment

Introduction

The use of pretrial detention has long been criticized for its differential impact on people of color (Ares, Rankin, & Sturtz, 1963; Petersen, 2020). There are further growing concerns about the influence of a person's sex in combination with their race-ethnicity on judicial decision-making during the pretrial release process. Pretrial scholarship routinely shows that racial-ethnic minorities are more likely to be detained prior to trial than White individuals (Demuth, 2003; Schlesinger, 2005). Research further suggests that while White females are the most likely to be released during pretrial, racial-ethnic minority males are the most likely to remain detained (Demuth, 2003; Schlesinger, 2005). One of the leading theoretical explanations for these disparities is that pretrial decisions are governed more by informal norms and organizational practices than by formal legal rules (Sutton, 2013). As judges and prosecutors receive incomplete information and maintain broad discretion during the pretrial process, they are believed to base their decisions in part on stereotypes about sex and racial-ethnic groups (Albonetti, 1991). Simply put, if males and people of color are perceived to pose a greater threat for failing to return to court and engaging in new criminal conduct, then judges and prosecutors may be more likely to take actions for detaining individuals with these characteristics regardless of the actual risk that each person may pose.

It is critical to ensure that pretrial processes are administered in a fair and equitable manner (Goldkamp, 1993; Van Brunt & Bowman, 2018). One of the most prominent strategies introduced as a potential mechanism for mitigating against the disparate treatment of people charged with crimes is the use of an actuarial pretrial assessment (Desmarais, Monahan, & Austin, 2022). The predictive validity of several pretrial instruments in identifying one's likelihood for new criminal activity while out on release has been established in numerous

jurisdictions across the United States (Desmarais, Zottola, Duhart Clarke, & Lowder, 2021). Nevertheless, support for pretrial assessments has been hindered by apprehensions that these tools might exacerbate sex and racial-ethnic disparities (Angwin, Larson, Mattu, & Kirchner, 2016; Eckhouse, Lum, Conti-Cook, & Ciccolini, 2019; Mayson, 2019; Woldgabreal, Day, & Tamatea 2020). The concern is that if biases exist in the scoring and prediction of an instrument, then implementing these tools within jurisdictions could serve to worsen the disparities observed in the criminal legal system. Recent research in pretrial assessments has largely focused on assessing scoring and prediction biases between sex *or* racial-ethnic groups separately and the findings show little evidence of bias (DeMichele, Baumgartner, Wegner, Barrick & Comfort, 2020; Desmarais et al., 2021). Although this scholarship is integral to understanding and developing strategies to reduce biases in pretrial assessments, the lack of research on the intersectionality of sex *and* race-ethnicity represents a substantial gap in the literature. This gap in knowledge is particularly concerning given that more than half of all pretrial agencies report using an assessment (Lattimore, Tueller, Levin-Rector, & Witwer, 2020).

In the current study, we seek to extend prior research by moving away from the "tendency to use a single-axis framework that treats race, ethnicity, and sex as mutually exclusive categories of experience and analysis" (Crenshaw, 1989, p. 139). More specifically, we assess for predictive bias in a widely used pretrial instrument—the *Public Safety Assessment (PSA)*—across six racial-ethnic and sex groups (i.e., White males, White females, Black males, Black females, Hispanic males, Hispanic females). This study was conducted as part of the Advancing Pretrial Policy and Research (APPR) project, which included the collection of historical administrative data from jail, court, and criminal history repositories in several counties across multiple states (see Arnold Ventures, 2022). The main statistical analyses

employed in this study involve assessing mean score differences, predictive validity, and differential prediction in the PSA across the six racial-ethnic and sex groups. The research and policy implications of the study findings are discussed.[1]

Pretrial Decisions

The public expects judges and prosecutors to make legal determinations that are free of bias and based on the strength of the evidence, quality of the investigation, and adherence to the law. Decisions about pretrial, however, must often be made quickly with little opportunity for these court officials to investigate or review evidence (Stevenson, 2018). Although judges and prosecutors receive information from multiple sources during the pretrial process, this evidence often varies in its content, amount, and complexity (DeMichele, Comfort, Misra, Barrick, & Baumgartner, 2021). Given the serious time constraints and workload pressures that are faced, judges and prosecutors must often make pretrial decisions about many people in a short amount of time based only on a quick review of the limited available information. There is also little formal oversight of the pretrial process itself, which raises concerns about fairness and equity in which individuals are released or detained prior to trial (Sutton, 2013).

Legal scholars have long debated the quality of judicial decision-making generally, with some who argue that such decisions are rooted in "hunches" or intuitive thinking (Frank, 1930). Psychological and behavioral economic researchers have further challenged traditional rational choice theories by showing humans often rely on cognitive shortcuts or heuristics in decision-making that can result in systematic errors in judgement (Kahneman, 2011; Stanovich & West, 2000). In the context of pretrial, there are concerns that time constraints and work pressures may

[1] The authors of the current paper were not involved in the development and validation research used to develop the PSA. The authors are currently engaged in several research projects on the PSA as part of a research partnership, but they are not personally invested in the PSA.

lead judges and prosecutors to miss or inconsistently weight relevant information when making decisions (DeMichele et al., 2021; Thaler, 2016). It is particularly alarming to consider that these courtroom actors may rely on cognitive heuristics, including stereotypes about criminal behavior and dangerousness to inform their decisions about pretrial release (Kang et al., 2011). If judges and prosecutors unwittingly associate males and racial-ethnic minorities with criminality, for instance, they may be more likely to take actions to detain rather than release individuals with these characteristics prior to trial, which stands in stark contrast to the public expectations that judicial officials should make decisions that are free from bias (Hochschild & Weaver, 2007; Schlesinger, 2005).

One strategy that has been advanced for its potential to reduce bias in pretrial decision-making is the use of actuarial assessments (Desmarais et al., 2022). According to Kleinberg et al. (2019), pretrial instruments can help to provide "clarity and transparency about the ingredients and motivations of [judicial] decisions" (p. 38). Assessments may further help judges and prosecutors reduce their reliance on intuitive judgements by providing a cognitive override to avoid heuristics (DeMichele et al., 2021). Despite their widespread adoption in pretrial jurisdictions across the United States (Lattimore et al., 2020), some scholars contend that instruments might inadvertently exacerbate racial-ethnic and sex disparities (Angwin et al., 2016; Eckhouse et al., 2019; Mayson, 2019; Woldgabreal et al., 2020).

Pretrial Assessments

Pretrial assessments commonly seek to integrate sets of items into a unified scale that can distinguish one's likelihood for negative pretrial outcomes if released (e.g., failure to appear in court, new criminal arrest, new criminal violent arrest; see Lowenkamp, 2009). In practice, this information can then be used to better inform pretrial release decisions and supervision plans

(Lowder et al., 2020). Empirical scholarship demonstrates that pretrial instruments are generally *good* to *excellent* predictors of these corresponding outcomes (Desmarais et al., 2021). Despite high levels of support for the predictive validity of pretrial assessments, the historic and ongoing differential treatment of males and people of color within the criminal legal system necessitate scrutiny of these tools through regular assessments of empirical validity and predictive biases (Eckhouse et al., 2019).

Although criminology has yet to agree on a conceptualization of what represents a biased assessment, standards in psychology, education, and organization studies often consider an instrument to be biased when evidence suggests that its "scores function differently for different groups" when making predictions (Vincent & Viljoen, 2020, p. 4). In the criminological literature, Jennifer Skeem, Christopher Lowenkamp, and John Monahan published two papers applying the definitions from the *Standards for Educational and Psychological Testing* (see American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) to evaluations of bias focused on a post-conviction assessment (i.e., the *Post Conviction Risk Assessment [PCRA]*). More specifically, Skeem, Monahan, and Lowenkamp (2016) compared predictive validity and tested calibration on the PCRA to determine if a given score on the scale had the same meaning for males and females. Similarly, Skeem and Lowenkamp (2016) assessed predictive accuracy and calibration for Black and White individuals on the PCRA as well.

These Standards have since been used by other scholars to assess for predictive bias in pretrial assessments and the findings have been mixed. For example, DeMichele et al. (2020) conducted a validation and bias test of the *Public Safety Assessment (PSA)* in Kentucky and found that two of the instrument's three scales —the New Criminal Activity (NCA) and New

Violent Criminal Activity (NVCA)—were equally predictive and calibrated across Black and White individuals. In contrast, the Failure to Appear (FTA) scale was found to be a significantly better predictor among White versus Black individuals ($p < .001$). More recently, a systematic review of the pretrial assessment literature noted that although predictive validity was generally found to be comparable across racial-ethnic groups, there remains little research to assess for differential prediction across other subgroups of interest (Desmarais et al., 2021). Of particular concern, there is a notable absence of scholarship assessing for predictive bias at the intersection of race-ethnicity and sex.

Current Study

In the criminal legal system, one's racial-ethnic heritage and biological sex are not treated independent of each other (Collier, 2019; Freiburger & Burke, 2011; Shuck et al., 2004; Stewart, 2007). Rather, these characteristics are believed to simultaneously contribute to disparities in criminal legal processing (Brennan & Spohn, 2009; Holmes et al., 2020). For example, a longstanding body of literature has suggested that Black and Hispanic males are treated more harshly by the American criminal legal system than White males (Brennan & Spohn, 2009; Holmes et al., 2020; Mauer, 1990). Evidence further suggests, however, that White males are treated more harshly by the criminal legal system than Black females, who are treated more harshly than White females (Brennan & Spohn, 2009; Holmes et al., 2020). The unequal application of punishment within the court system appears to extend to the pretrial system, where Black males tend to have the highest likelihood of being detained during pretrial (Sacks et al., 2015; Spohn, 2008). Moreover, research suggests that White females face the lowest likelihood of being detained during pretrial, with some scholars positing that this association is due in part to the perception of their frailness by court officials (Hood & Schneider, 2019).

Although the mechanisms underlying unequal treatment by the pretrial system is open for debate, the longstanding disparities in the use of detention underscore the need for pretrial assessments that are not biased across racial-ethnic and sex groupings. Scholarship, however, has been limited to assessments of racial-ethnic *or* biological sex differences in the predictive validity and biases. To date, no research has examined the existence of differential prediction in pretrial assessments across racial-ethnic *and* sex groupings. In response, this study involves a multi-site validation and bias test of the PSA across six groups: White males, White females, Black males, Black females, Hispanic males, and Hispanic females. More specifically, we address three research questions:

1. **Descriptive Comparison.** *Does the PSA have similar scores across race-ethnicity and sex groups.*
2. **Predictive Validity**. *Does the PSA have a similar predictive accuracy across race-ethnicity and sex groups?*
3. **Differential Prediction**. *Does the PSA have scales that perform equally well across race-ethnicity and sex groups?*

If we do not find significant differences for differential prediction, our analyses would support the tools as equitable across race-sex groups; however, if we find significant differences between the groups, that would call into question the use of the tools.

Method

Sample

The data for the current study was collected as part of the Advancing Pretrial Policy and Research (APPR) project (see Arnold Ventures, 2022). To date, APPR has involved the provision of technical and research assistance across three jurisdictions in two states. The larger

goal of this project has been to conduct and disseminate research that can help improve the pretrial system and outcomes of justice involved individuals. In the current study, we focus on data obtained from three counties in two states, including two counties from a northwestern state and one county from a southeastern state. These three counties were selected because they provided sufficient jail, court, and criminal history information to retrospectively score the PSA. Only people who were booked and released to the community prior to the disposition of their case were included in our analytic sample, as those who were detained for the entire pretrial period were never at risk of experiencing a pretrial outcome. Additionally, our analytic sample was limited to those whose racial-ethnic heritage was listed as *White*, *Black*, or *Hispanic* in the official data, as individuals with other racial-ethnic heritages were removed due to their small sample size. In total, our inclusion criteria resulted in 26,336 unique individuals who were booked and released between January 2017 and December 2018.

Measures

Public Safety Assessment

The Public Safety Assessment (PSA) is an actuarial pretrial instrument that uses criminal history and current offense information to construct separate scales that predict the likelihood of three outcomes from occurring during pretrial release, including failure to appear (FTA), new criminal arrest (NCA), and new violent criminal arrest (NVCA).[2] The PSA relies on nine items: (1) age at current arrest, (2a) current violent offense, (2b) current violent offense and 20 years old or younger, (3) pending charge at the time of arrest, (4) prior misdemeanor conviction, (5a) prior felony conviction, (5b) prior conviction (misdemeanor or felony), (6) prior violent

[2] For more information on the development, implementation, and use of the PSA, see https://advancingpretrial.org/psa/factors/.

conviction, (7) prior failure to appear in the past two-years, (8) prior failure to appear older than two-years, and (9) prior sentence to incarceration. The FTA, NCA, and NVCA scales include different compositions of items which have been weighted by the strength of their relationship with each respective pretrial outcome. The FTA scale uses four of the nine items (3, 5b, 7, and 8), the NCA scale uses seven items (1, 3, 4, 5a, 6, 7, and 9), and the NVCA scale uses five items (2a, 2b, 3, 5b, and 6). These scores are converted to range in value from 1 to 6, with lower scores indicating greater likelihood of pretrial success and higher scores indicated greater likelihood of pretrial failure. Research on the PSA has consistently demonstrated that these three scales are strong and valid predictors of the corresponding outcomes (DeMichele et al., 2020; DeMichele & Baumgartner, 2021; Griener et al., 2020; Lowenkamp, DeMichele, & Klein-Warren, 2020).

In the three research sites, the PSA was not implemented between January 2017 and December 2018. As such, court officials did not have access to PSA scores to inform their release decisions. In this study, we retrospectively calculated the PSA scores for the FTA, NCA, and NVCA scales from the administrative records collected from the three counties using the psa2013 package in R (Tueller et al., 2022). Such retrospective calculations are possible because the PSA relies only age and criminal history information that does not require an in-person interview.

Race-Ethnicity and Sex Groupings

Race and sex information was self-reported to corrections officials during jail intake process. This information was used to create the race-ethnicity and sex groupings. For one northwestern county, we were able to construct six race-ethnicity and sex groups, including White males, White females, Black males, Black females, Hispanic males, and Hispanic females.

In the other two counties, the analyses are limited to White and Black male and female groups because the data provided did not include an indicator for ethnic heritage.

Outcomes

This study includes three dichotomous pretrial outcomes. First, *failure to appear (FTA)* measures whether someone missed a court date. Second, *new criminal arrest (NCA)* measures whether someone was arrested for a new crime. Third, *new violent criminal arrest (NVCA)* measures whether someone was arrested for a new violent crime. These outcomes were all constructed across one's pretrial period from the administrative records. For those who did not experience a disposition during the observation period, their outcome data was collected through December 2019.[3]

Analytical Strategy

We examine and compare descriptive statistics across the three county samples, including the average PSA scale scores and pretrial outcomes by racial-ethnic and sex group. Next, we evaluate the predictive validity for the PSA by using Area Under the Curve (AUC) Receiver Operator Characteristics (ROC) estimates. The AUC ranges from 0 to 1.0 with .5 referring to no better than random chance and 1.0 referring to perfect prediction. The AUC has an intuitive interpretation by reporting the likelihood that when randomly selecting a case that had one of the outcomes, that case would have a higher score on the PSA than a randomly selected case that did not have one of the outcomes. According to Desmarais, Johnson, and Singh (2018), AUC values

[3] There were 4,727 individuals who did not have disposition on their case by December 31, 2019. Supplemental analyses were conducted to assess if censorship may have influenced the validation and bias evaluation of the PSA. The sensitivity analyses showed that the censorship had limited influence on the PSA validation and bias tests. In instances where the results differed, the predictive validity of the PSA increased due to reducing the number of non-recidivists in the model. The results of the models evaluating the predictive bias by race-ethnicity and sex groupings supported the conclusions of the main findings. Further, parametric survival models also supported the main findings reported here.

of .54 and below are defined as *poor* predictors, .55 to .63 are *fair* predictors, and .64 to .70 are *good* predictors, and values higher than .710 are *excellent* predictors. AUCs values between groups were compared using the Venkatraman method test for evaluating if the receiver operating characteristic curve (ROC) for one group can be superimposed on the ROC curve for another curve (Venkatraman, 2000).

We estimate a series of logistic regression models to assess for evidence of predictive bias by race-ethnicity and sex with the PSA. More specifically, we conduct four regression models for each outcome for each county to assess the extent to which "a given score will have the same meaning regardless of group membership" (Monahan, Skeem, & Lowenkamp, 2017, p. 193). Differential prediction is tested by assessing the extent to which subgroups have similar (i.e., not statistically significantly different) intercepts and slopes (i.e., they possess similar regression lines). The approach is designed to test for whether the PSA scores are moderated or conditioned by race-ethnicity and sex group as they predict the outcomes (i.e., a given score on the PSA does not have the same meaning for White males and Black males). Because of the large sample size and the number of comparisons made in this study, a more stringent definition of statistical significance was used for interpreting the results (i.e., $p < .001$ vs. .05; McShane et al., 2019; Skeem and Lowenkamp, 2016).

Results

Descriptive Comparisons

Table 1 compares the PSA scores and pretrial outcomes across the six racial-ethnic and sex groups in the three counties. As can be seen in the table, Black males in County 1 had the highest FTA and NCA scores, as well as the highest rates of new criminal arrest and new violent criminal arrest, followed by White males, White females, and Black females. These findings,

however, were not observed for counties 2 and 3, where Black males scored higher than White males, White females, and Black females on the NCA. Black and White males had a higher score on the NCA in County 2 and on the NVCA in County 3. Regarding the outcomes of interest, limited differences were found between the race-ethnicity and sex groupings on the rates of failure to appear, new criminal arrest, and new violent criminal arrest.

*** Insert Table 1 About Here ***

Predictive Validity

Table 2 presents the AUCs for the FTA, NCA, and NVCA scales across the race-ethnicity and sex groupings in the three counties. Across the 42 comparisons, there was only one instance detected of a statistically significant difference in AUC values between two race-ethnic and sex groups. This difference was in County 1 with the FTA scale in which Black males had a significantly lower AUC than White females. The reported AUCs for the race-ethnicity and sex groupings were *fair* to *good*, except for the AUC for White females in County 2 which was categorized as *poor*.

*** Insert Table 2 About Here ***

Differential Prediction

Table 3 presents the logistic regression models assessing predictive bias of the FTA scale for the three counties. White males were selected as the reference group because we were interested in the performance of the PSA for the other race-ethnicity and sex groupings. First, we find that a one-unit increase in the FTA scale is associated with significant increases ($p < .001$) in the odds of an FTA by 33%, 14%, and 48% for counties 1 through 3, respectively. Second, we find one significant ($p < .001$) direct effect with Black males ($OR = 1.32$) in County 1 (model 2) associated with FTAs. Third, the effect of race-ethnicity and sex on FTAs does not exist in

model 3 when we add the FTA scale, with the latter variable remaining significant (*OR* = 1.33, *p* < .001). Fourth, the most important concern when testing for moderation effects is whether any of the race-ethnicity and sex by score interactions are significant. None of the interaction terms are found to be significant for the FTA scale, indicating that the FTA scale predicts recidivism similarly across race-ethnicity and sex groupings. This finding was further supported by the estimation of a series of equality of coefficients tests – a Wald-test-based comparison between a model and a linearly restricted model – to evaluate if the predictive validity of the FTA differed across any of the race-ethnicity and sex groupings (Feltz and Miller, 1996). The results of these tests suggested that no statistically significant differences existed in the predictive validity of the FTA across the race-ethnicity and sex groupings.

*** Insert Table 3 About Here ***

In Table 4, we report the results of the logistic regression models to assess predictive bias for the NCA scale. We find nearly identical results for NCA as we found for FTA. First, we find a one-unit increase in the NCA scale is associated with a 50%, 28%, and 49% increase in the odds of an NCA for counties 1 through 3, respectively. Second, we find direct effects for Black males in County 1 (model 2, *OR* = 1.44, *p* < 0.001), but these effects do not remain in model 3 when adding NCA scale score to the model (*OR* = 1.13, *p* = .061). Lastly, we do find that the Black males interaction term is significant, with an odds ratio (0.82; *p* < .001) indicating that the NCA scale does not predict new arrests for Black males as well as White males. The probability of recidivism associated with higher scores on the NCA scale for Black males is lower than the probability of recidivism associated with higher scores on the NCA scale for White males, while the probability of recidivism associated lower scores on the NCA scale for Black males is higher

than the probability of recidivism associated lower scores on the NCA scale for White males. This moderation was not observed in counties 2 or 3.

*** Insert Table 4 About Here ***

In Table 5, we test for predictive bias with the NVCA scale. We find that the NVCA scale is positively associated with new violent arrests. In particular, the results suggested that a 1-point increase in the NVCA scale was associated with a 62% increase in the odds of recidivism for County 1 ($p < .001$), a 75% increase in the odds of recidivism for County 2 ($p < .001$), and a 60% increase in the odds of recidivism for County 3 ($p < .001$). Being a Black male in County 1 did appear to be associated with a 63% increase in the odds of recidivism ($p < .001$), an effect not observed for counties 2 and 3. Moreover, the effects of race-ethnicity and sex groupings on new violent arrests were not observed when the model accounted for the NVCA score. Regarding model 4, the results presented in Table 5 suggest that the race-ethnicity and sex groupings do not moderate the predictive validity of the NVCA scale. Specifically, the race-ethnicity and sex groupings did not appear to increase or decrease the predictive validity of the NVCA scale in any of the three counties.

*** Insert Table 5 About Here ***

Discussion

Over the past two decades, pretrial assessments have been adopted in jurisdictions across the United States as a strategy for providing courtroom actors with more objective information about an individual's propensity toward criminal behavior if released prior to his or her trial (Lattimore et al., 2020). Prior research has provided overwhelming support for the ability of these instruments in predicting pretrial outcomes of interest, including failure to appear in court and new criminal arrests (Desmarais et al., 2021). Although some champion pretrial assessments

for their ability to increase fairness and transparency in pretrial decision-making processes, others express concerns that these tools might serve to exacerbate the sex and racial-ethnic disparities found in the larger criminal justice system (Eckhouse et al., 2019; Mayson, 2019; Woldgabreal et al., 2020). In response to this debate, a small, but growing, body of scholarship has sought to assess if these instruments possess predictive biases across sex *or* racial-ethnic groups and the findings have been consistently found little evidence for predictive bias (DeMichele et al., 2020; Desmarais et al., 2021; Lowder et al., 2022).

In the current study, we expand on this knowledge base by conducting the first known predictive bias analysis of a pretrial assessment at the intersectionality of sex *and* race-ethnicity. More specifically, this multi-site validation study used data collected as part of the APPR project to test for evidence of predictive bias and differential prediction in the PSA across six groups, including White males, White females, Black males, Black females, Hispanic males, and Hispanic females. Given the concerns that a defendant's sex and race-ethnicity may have on pretrial decisions, the inclusion and analysis of this combination of ascriptive characteristics is of significant theoretical and practical importance. Toward this end, this investigation yielded three findings that have important implications for research and policy.

First, the descriptive comparisons of the PSA scales indicated several meaningful differences in the mean scores across the six sex and race-ethnic groups examined. In County 1, for example, Black males scored the highest on the three PSA scales, followed by Black females, White males, and White females. In County 2, Black females scored the highest on the FTA and NCA scales, while Black males scored the highest on the NVCA scale. In County 3, Black males scored the highest on the three scales, followed by White males, Black females, White females, Hispanic males, and Hispanic females. These observed differences in scores, however, should be

interpreted cautiously. The sex and race-ethnic group with the highest scale score in each county also had the highest rate of the corresponding outcome.[4] This finding suggests that higher scores across some groupings is not necessarily evidence of bias, but rather may simply reflect real differences in the likelihood of the three pretrial outcomes across the sex and race-ethnic groups.

Second, the three PSA scales were found to be *fair* to *good* predictors of failure to appear, new criminal arrest, and new violent criminal arrest across the six sex and racial-ethnic groups examined in the three counties. The comparison of AUC values across the three PSA scales revealed evidence of only one statistically significant difference between two separate sex and racial-ethnic groups. More specifically, the FTA scale was found to be a significantly better predictor of failure to appear events in County 1 for White females (AUC = .685) relative to Black males (AUC = .603). It should be noted, however, that while the AUC values were significantly lower in this county for Black males, the FTA scale was nonetheless still a fair predictor for this sex and racial-ethnic group. Taken together, these findings provide support for the predictive validity of the three PSA scales across these six sex and race-ethnicity groups.

Third, the analyses showed a lack of evidence suggesting differential prediction for the three PSA tools across the six racial-ethnic and sex groupings. The regression models failed to produce evidence of predictive bias when testing for moderation. We estimated 36 logistic regression models and found one statistically significant interaction term in which Black males were under predicted to have a new arrest (relative to White males) in County 1. Our findings

[4] There were a couple of exceptions to this statement. In County 2, for example, while Black females had the highest NCA score, White females had the highest observed rates of new criminal activity (21% versus 20%, respectively). Additionally, while Black males had the highest NVCA score, all four sex and race-ethnicity groups shared similar observed rates of new violent criminal activity (5%). In County 3, while Black males had the highest FTA and NVCA scores, Hispanic females had the highest observed rate of failure to appear (36% versus 27%, respectively) and shared the same observed rate of new violent criminal activity (4%).

align with and extend prior research on predictive bias with pretrial assessments (Desmarais et al., 2022; Vincent & Viljoen 2020). The lack of evidence of differential validity or prediction across so many tests provides support for the empirical foundation and the implementation of pretrial assessments.

Limitations and Future Directions

Although this study informs our understanding of pretrial assessments, there are methodological limitations that should be understood when interpreting its results. First, the current study relied on data collected from three county agencies in two states that volunteered to take part in the APPR project. It is possible that these jurisdictions possess unique characteristics, such as the willingness to participate in a research project aimed at reforming pretrial practices, which may have had an influence on the study findings. We therefore caution against generalizing these results to other counties at this stage and encourage the continued study of predictive biases in pretrial assessments across a more diverse set of jurisdictions.

Second, this study focused its analysis on three categories of race-ethnicity (White, Black, and Hispanic). Due to data constraints and the demographic composition of the samples from these counties, we were only able to include Hispanic individuals in one of the three research sites (i.e., County 3). Study findings, therefore, should be interpreted cautiously for Hispanic individuals and not be generalized to other race-ethnicity groups. Future research should include larger samples of Hispanic individuals, as well as involve other racial-ethnic groups (e.g., Asian and Native American individuals). In addition, there are other potential demographic characteristics that may be theoretical importance that this study was unable to investigate due to its sample size. The inclusion of age, for example, might help provide more informative subgroup (i.e., age, sex, *and* race-ethnicity groupings). With the addition of each

attribute, there is a requisite need to ensure that there are enough cases in each grouping to conduct the statistical analyses.

Finally, the current study involved the analysis of administrative court data, jail records, and criminal history records. Although the data collected was sufficient to retrospectively calculate the three PSA scale scores, this information was not available to the judges and prosecutors at the time they had to make their initial pretrial decisions. Future scholarship should seek to include prospective investigations of predictive biases in pretrial instruments as used in practice.

Conclusion

Amidst concerns that the use of assessments may further contribute to the disparities found in the criminal legal system (Eckhouse et al., 2019; Mayson, 2019; Woldgabreal et al., 2020), this study supports the PSA as a valid predictor of pretrial outcomes across the six sex and racial-ethnic groups examined (i.e., White males, White females, Black males, Black females, Hispanic males, Hispanic females). According to the study findings, the PSA has the potential to provide judges and prosecutors with valuable empirical information that can be used to make informed decisions about whether and under what conditions to release someone prior to trial (Lowder et al., 2020). The use of assessment information holds promise for alleviating some of the potential harms done by unnecessarily detaining people who are at low risk for failing to appear or engaging in new criminal activity (Campbell et al., 2020; Goulette & Wooldredge, 2018). Despite these positive results, it is critical that researchers continue to engage in strategies to develop and refine pretrial instruments in ways that can further improve predictive accuracy and mitigate against potential biases. The implementation of pretrial assessments is not the answer to all the problems that plague pretrial systems, but they do hold the potential to provide an

empirical basis for improving such systems. As such, we encourage the continued study of pretrial assessments.

References

Albonetti, C. A. (1991). An integration of theories to explain judicial discretion. *Social problems*, *38*(2), 247-266.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association. https://www.apa.org/science/programs/testing/standards

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals, and its biased against Blacks. *ProPublica*. https://www.propublica.org/article/machine-biasrisk-assessments-in-criminal-sentencing.

Ares, C.E., Rankin, A. and Sturtz, H. (1963) 'The Manhattan bail project: an interim report on the use of pretrial parole', *New York University Law Review, 38*, 67–95.

Arnold Ventures. (2022). *About the Public Safety Assessment*. Retrieved from https://advancingpretrial.org/psa/about

Brennan, P. K., & Spohn, C. (2009). The joint effects of offender race/ethnicity and sex on sentence length decisions in federal courts. *Race and Social Problems*, *1*(4), 200-217.

Campbell, C. M., Labrecque, R. M., Weinerman, M., & Sanchagrin, K. (2020). Gauging detention dosage: Assessing the impact of pretrial detention on sentencing outcomes using propensity score modeling. *Journal of Criminal Justice*, *70*, 101719.

Collier, C., S (2019). Black males and Courts (pp. 49-58). In Williams, J. M., & Kniffley, S. (Eds.) *Black males and the criminal justice system*. Routledge.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. *The University of Chicago Legal Forum*, (1), 139-167.

DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020). Public safety assessment: Predictive utility and differential prediction by race in Kentucky. *Criminology & Public Policy*, *19*(2), 409-431.

DeMichele, M., Comfort, M., Barrick, K., & Baumgartner, P. (2021). The intuitive-override model: Nudging judges toward pretrial risk assessment instruments. *Federal Probation*, *85*, 22-39.

Demuth, S. (2003). Racial and ethnic differences in pretrial release decisions and outcomes: A comparison of Hispanic, Black, and White felony arrestees. *Criminology*, *41*(3), 873-908.

Desmarais, S. L., Monahan, J., & Austin, J. (2022). The empirical case for pretrial risk assessment instruments. *Criminal Justice and Behavior*, *49*(6), 807-816.

Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2018). Performance of recidivism risk assessment instruments in US correctional settings (pp. 1-29). In J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, & Z. Hamilton (Eds.). *Handbook of recidivism risk/needs assessment tools*. John Wiley & Sons.

Desmarais, S. L., Zottola, S. A., Duhart Clarke, S. E., & Lowder, E. M. (2021). Predictive validity of pretrial risk assessments: A systematic review of the literature. *Criminal Justice and Behavior*, *48*(4), 398-420.

Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, *46*(2), 185-209.

Frank, J. (1930). *Law and the modern mind*. Coward-Mccann.
Feltz, C. J., & Miller, G. E. (1996). An asymptotic test for the equality of coefficients of variation from k populations. *Statistics in medicine*, *15*(6), 647-658.
Freiburger, T. L., & Burke, A. S. (2011). Status offenders in the juvenile court: The effects of gender, race, and ethnicity on the adjudication decision. *Youth Violence and Juvenile Justice*, *9*(4), 352-365.
Goulette, N., & Wooldredge, J. (2018). Collateral consequences of pretrial detention (pp. 271-289). In B. M. Huebner, & N. A. Frost (Eds.) *Handbook on the Consequences of Sentencing and Punishment Decisions*. Routledge.
Goldkamp, J. S., & Weiland, D. (1993). *Assessing the impact of Dade County's felony drug court*. National Institute of Justice.
Griener, D. J., Stubenberg, M., Halen, R. (2020) Validation of the PSA in Harris County, TX. Technical Report Submitted to *Advancing Pretrial Policy and Research, Arnold Ventures.*
Holmes, B., Feldmeyer, B., & Kulig, T. C. (2020). Sentencing departures and focal concerns: The joint effect of race and gender on departures in United States district courts, 2014–2016. *Journal of Crime and Justice*, *43*(5), 598-622.
Hood, K., & Schneider, D. (2019). Bail and pretrial detention: Contours and causes of temporal and county variation. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(1), 126-149.
Hochschild, J. L., & Weaver, V. (2007). The skin color paradox and the American racial order. *Social Forces*, *86*(2), 643-670.
Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
Kang, J., Bennett, M., Carbado, D., Casey, P., & Levinson, J. (2011). Implicit bias in the courtroom. *UCLA Law Review*, *59*, 1124.
Kleinberg, J., Ludwig, J., Mullainathan, S., & Sustein, C. R. (2019). *Discrimination in the age of algorithms*. Arxiv. https://arxiv.org/abs/1902.03731
Lattimore, P. K., Tueller, S., Levin-Rector, A., & Witwer, A. (2020). The prevalence of local criminal justice practices. *Federal Probation*, *84*, 28-37.
Lowder, E. M., Lawson, S. G., Grommon, E., & Ray, B. R. (2020). Five-county validation of the Indiana Risk Assessment System–Pretrial Assessment Tool (IRAS-PAT) using a local validation approach. *Justice Quarterly*, *37*(7), 1241-1260.
Lowder, E. M., Diaz, C. L., Grommon, E., & Ray, B. R. (2022). Differential prediction and disparate impact of pretrial risk assessments in practice: a multi-site evaluation. *Journal of Experimental Criminology*, advanced online publication.
Lowenkamp, C., DeMichele, M., & Klein Warren, L. (2020). *Replication and extension of the Lucas County PSA Project*. Technical Report Submitted to *Advancing Pretrial Policy and Research, Arnold Ventures.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727443.
Mauer, M. (1990). Young Black men and the criminal justice system. *Corrections Compendium*, *15*(2), 1-4.
Mayson, S. G. (2019). Detention by Any Other Name. *Duke Law Journal*, *69*, 1643-1680.
McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(sup1), 235-245.

Monahan, J., Skeem, J., & Lowenkamp, C. (2017). Age, risk assessment, and sanctioning: Overestimating the old, underestimating the young. *Law and Human Behavior*, *41*(2), 191.
Petersen, N. (2020). Do detainees plead guilty faster? A survival analysis of pretrial detention and the timing of guilty pleas. *Criminal Justice Policy Review*, *31*(7), 1015-1035.
Sacks, M., Sainato, V. A., & Ackerman, A. R. (2015). Sentenced to pretrial detention: A study of bail decisions and outcomes. *American Journal of Criminal Justice*, *40*(3), 661-681.
Schlesinger, T. (2005). Racial and ethnic disparity in pretrial criminal processing. *Justice Quarterly*, *22*(2), 170-192.
Schuck, A. M., Lersch, K. M., & Verrill, S. W. (2004). The "invisible" Hispanic? The representation of Hispanics in criminal justice research: What do we know and where should we go?. *Journal of Ethnicity in Criminal Justice*, *2*(3), 5-22.
Skeem, J. L., and Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias an disparate impact. *Criminology*, 54, 680-712.
Skeem, J., Monahan, J., & Lowenkamp, C. (2016). Gender, risk assessment, and sanctioning: The cost of treating women like men. *Law and Human Behavior*, *40*(5), 580-593.
Spohn, C. (2008). Race, sex, and pretrial detention in federal court: Indirect effects and cumulative disadvantage. *University of Kansas Law Review. 57*, 879-901.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and Brain Sciences*, *23*(5), 645-665.
Stevenson, M. (2018). Assessing risk assessment in action. *Minnesota Law Review*, *103*, 303-384.
Stewart, E. A. (2007). Either they don't know or they don't care: Black males and negative police experiences. *Criminology & Public Policy*, *6*, 123-130.
Sutton, J. R. (2013). Structural bias in the sentencing of felony defendants. *Social Science Research*, *42*(5), 1207-1221.
Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American Economic Review*, *106*(7), 1577-1600.
Tueller S.J., Chew R., Godwin A.K., Williams J.A., & Weitzel K. (2022). _psa2013: Validation tools for the Public Safety Assessment (PSA)_. R package version 0.3.9, <https://osf.io/3gsmf/>.
Van Brunt, A., & Bowman, L. E. (2018). Toward a just model of pretrial release: a history of bail reform and a prescription for what's next. *Journal of Criminal Law & Criminology*, *108*, 701-774.
Venkatraman, E. S. (2000). A permutation test to compare receiver operating characteristic curves. *Biometrics*, *56*(4), 1134-1138.
Vincent, G. M., & Viljoen, J. L. (2020). Racist algorithms or systemic problems? Risk assessments and racial disparities. *Criminal Justice and Behavior*, *47*(12), 1576-1584.
Woldgabreal, Y., Day, A., & Tamatea, A. (2020). Do risk assessments play a role in the enduring 'color line'. *Advancing Corrections*, *10*, 18-28.

Tables and Figures

Table 1

Descriptive Comparisons of PSA Scores and Pretrial Outcomes across the Race-Ethnicity and Sex Groupings, by County

	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Anova	Group Difference ($p < .001$)
	White Males		White Females		Black Males		Black Females		Hispanic Males		Hispanic Females			
County 1														
Public Safety Assessment														
FTA Score	2.07	1.26	2.02	1.29	2.60	1.41	2.17	1.41	--	--	--	--	192.55*	b,f,j
NCA Score	2.10	1.13	1.92	1.04	2.74	1.37	2.06	1.14	--	--	--	--	455.81*	b,f,j
NVCA Score	1.71	.90	1.57	.82	1.97	1.08	1.74	.94	--	--	--	--	104.49*	b,f,g,j
Outcomes														
Failure to Appear	.14	.35	.14	.34	.18	.39	.13	.34	--	--	--	--	22.48*	b,j
New Criminal Arrest	.21	.41	.18	.39	.28	.45	.19	.39	--	--	--	--	64.61*	b,f,j
New Violent Criminal Arrest	.05	.23	.03	.17	.09	.28	.06	.24	--	--	--	--	22.72*	b,f,j
N	2,249		920		12,393		4,384		--		--			
County 2														
Public Safety Assessment														
FTA Score	3.17	1.53	3.22	1.62	3.04	1.50	3.42	1.62	--	--	--	--	3.61	
NCA Score	3.04	1.49	2.81	1.45	3.16	1.44	3.16	1.58	--	--	--	--	8.48*	a,f
NVCA Score	1.86	.96	1.61	.87	2.05	1.04	1.86	.98	--	--	--	--	27.43*	a,b,f
Outcomes														
Failure to Appear	.32	.47	.34	.47	.32	.46	.36	.48	--	--	--	--	.74	
New Criminal Arrest	.30	.46	.27	.44	.29	.45	.27	.45	--	--	--	--	.99	
New Violent Criminal Arrest	.08	.28	.06	.24	.09	.29	.07	.26	--	--	--	--	1.38	
N	1,905		818		717		197		--		--			
County 3														
Public Safety Assessment														
FTA Score	3.56	1.61	3.46	1.72	3.63	1.65	3.56	1.77	3.07	1.60	2.95	1.78	3.88	
NCA Score	3.43	1.60	3.15	1.62	3.58	1.68	3.39	1.73	3.04	1.61	3.03	1.60	5.38*	
NVCA Score	2.04	1.05	1.80	.93	2.12	1.10	1.89	.88	1.93	1.00	1.77	.90	6.62*	a,f
Outcomes														
Failure to Appear	.25	.43	.25	.43	.27	.44	.18	.39	.18	.38	.36	.49	1.85	
New Criminal Arrest	.17	.37	.15	.35	.21	.41	.15	.36	.17	.38	.21	.41	1.13	
New Violent Criminal Arrest	.04	.20	.03	.18	.07	.25	.00	.00	.04	.20	.05	.22	1.49	
N	1,573		689		226		61		165		39			

Notes: "a" = White Females different from White Males; "b" = Black Males different from White Males; "c" = Black Females different from White Males; "d" = Hispanic Males different from White Males; "e" = Hispanic Females different from White Males; "f" = Black Males different from White Females; "g" = Black Females different from White Females; "h" = Hispanic Males different from White Females; "i" = Hispanic Females different from White Females; "j" = Black Females different from Black Males; "k" = Hispanic Males different from Black Males; "l" = Hispanic Females different from Black Males; "m" = Hispanic Males different from Black Females; "n" = Hispanic Females different from Black Females; ; "n" = Hispanic Females different from Hispanic Males. For mean differences, the group difference column represents the statistical significance of a Tukey test. Tukey test evaluates which group means are statistically different from each other at a $p < .001$ level.

Table 2

Predictive Validity of the PSA across Race-Ethnicity and Sex Groupings, by County

	White Males	White Females	Black Males	Black Females	Hispanic Males	Hispanic Females	Overall AUC
				County 1			
FTA AUC	.652	.685	.603	.633			.622
NCA AUC	.676	.664	.633	.648			.653
NVCA AUC	.639	.727	.642	.641			.620
N	2,249	920	12,393	4,384	--	--	
				County 2			
FTA AUC	.569	.500	.570	.594			.555
NCA AUC	.598	.602	.601	.659			.603
NVCA AUC	.663	.691	.610	.656			.661
N	1,905	818	717	197	--	--	
				County 3			
FTA AUC	.694	.635	.635	.636	.671	.706	.672
NCA AUC	.677	.630	.695	.826	.682	.790	.674
NVCA AUC	.659	.578	.597	--	.736	.561	.642
N	1,573	689	226	61	165	39	

Notes: FTA AUC for Black Males was statistically different from FTA AUC for White Females in County 1. A bootstrapped test was used to evaluate if the ROC curves were statistically different across the groups at a $p < .001$ level.

Table 3

Differential Prediction of the FTA Tool, by Race-Ethnicity and Sex Grouping

DV: Failure to Appear	Model 1				Model 2				Model 3				Model 4			
	OR	95%CI		*p-value*	OR	95%CI		*p-value*	OR	95%CI		*p-value*	OR	95%CI		*p-value*
County 1																
FTA Score	**1.338**	**1.304**	**1.373**	**< .001**	--	--	--	--	**1.327**	**1.293**	**1.361**	**< .001**	**1.434**	**1.317**	**1.561**	**< .001**
White Females	--	--	--	--	.943	.753	1.174	.603	.953	.760	1.190	.675	.844	.532	1.320	.462
Black Males	--	--	--	--	**1.317**	**1.163**	**1.497**	**< .001**	1.128	.993	1.284	.066	1.482	1.142	1.936	.003
Black Females	--	--	--	--	.920	.796	1.066	.266	.883	.762	1.024	.098	.978	.727	1.322	.887
Interaction Terms																
WF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.051	.899	1.228	.535
BM*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.899	.821	.985	.022
BF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.956	.863	1.059	.391
N	19,946															
County 2																
FTA Score	**1.144**	**1.093**	**1.197**	**< .001**	--	--	--	--	**1.142**	**1.092**	**1.195**	**< .001**	**1.186**	**1.113**	**1.265**	**< .001**
White Females	--	--	--	--	1.080	.907	1.284	.386	1.072	.900	1.276	.433	1.841	1.236	2.736	.003
Black Males	--	--	--	--	.965	.802	1.160	.709	.981	.815	1.180	.843	.956	.613	1.479	.840
Black Females	--	--	--	--	1.182	.867	1.601	.285	1.143	.837	1.551	.394	.993	.453	2.091	.986
Interaction Terms																
WF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.848	.760	.946	.003
BM*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.010	.892	1.144	.876
BF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.038	.855	1.267	.713
N	3,638															
County 3																
FTA Score	**1.484**	**1.400**	**1.575**	**< .001**	--	--	--	--	**1.486**	**1.401**	**1.578**	**< .001**	**1.598**	**1.473**	**1.737**	**< .001**
White Females	--	--	--	--	.986	.800	1.212	.893	1.007	.811	1.248	.948	2.046	1.134	3.651	.016
Black Males	--	--	--	--	1.121	.813	1.529	.476	1.093	.784	1.509	.592	2.091	.805	5.051	.114
Black Females	--	--	--	--	.667	.327	1.246	.232	.636	.306	1.215	.194	1.248	.147	6.566	.815
Hispanic Males	--	--	--	--	.647	.419	.967	.041	.764	.489	1.159	.221	1.060	.320	3.053	.919
Hispanic Females	--	--	--	--	1.699	.852	3.255	.118	2.247	1.088	4.480	.024	2.929	.545	12.194	.167
Interaction Terms																
WF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.838	.732	.962	.011
BM*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.853	.693	1.061	.142
BF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.850	.571	1.340	.447
HM*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.922	.703	1.230	.568
HF*FTA Score	--	--	--	--	--	--	--	--	--	--	--	--	.939	.632	1.462	.765
N	2,753															

Notes: Estimates with $p < .001$ are bolded in the table. "WF" = White Female; "BM" = Black Male; "BF" = Black Female; "HM" = Hispanic Male; "HF" = Hispanic Female.

Table 4

Differential Prediction for NCA Tool, by Race-Ethnicity and Sex Grouping

DV: New Criminal Arrest	Model 1				Model 2				Model 3				Model 4			
	OR	95%CI		*p-value*	OR	95%CI		*p-value*	OR	95%CI		*p-value*	OR	95%CI		*p-value*
County 1																
NCA Score	**1.503**	**1.467**	**1.540**	**< .001**	--	--	--	--	**1.476**	**1.440**	**1.514**	**< .001**	**1.731**	**1.587**	**1.890**	**< .001**
White Females	--	--	--	--	.842	.692	1.021	.084	.905	.741	1.101	.321	1.022	.651	1.589	.923
Black Males	--	--	--	--	**1.441**	**1.294**	**1.607**	**< .001**	1.113	.996	1.246	.061	**1.853**	**1.438**	**2.401**	**< .001**
Black Females	--	--	--	--	.848	.748	.963	.011	.854	.751	.972	.016	1.016	.759	1.363	.918
Interaction Terms																
WF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.960	.807	1.142	.641
BM*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	**.818**	**.745**	**.896**	**< .001**
BF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.931	.835	1.036	.189
N	19,946															
County 2																
NCA Score	**1.283**	**1.222**	**1.348**	**< .001**	--	--	--	--	**1.283**	**1.222**	**1.349**	**< .001**	**1.264**	**1.183**	**1.352**	**< .001**
White Females	--	--	--	--	.862	.717	1.033	.110	.911	.756	1.096	.327	.855	.553	1.311	.475
Black Males	--	--	--	--	.926	.765	1.118	.427	.899	.741	1.088	.277	.845	.517	1.365	.495
Black Females	--	--	--	--	.873	.625	1.205	.419	.842	.599	1.168	.311	.535	.220	1.212	.149
Interaction Terms																
WF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.020	.900	1.158	.754
BM*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.019	.893	1.165	.780
BF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.140	.917	1.432	.247
N	3,638															
County 3																
NCA Score	**1.486**	**1.390**	**1.591**	**< .001**	--	--	--	--	**1.486**	**1.389**	**1.592**	**< .001**	**1.508**	**1.377**	**1.655**	**< .001**
White Females	--	--	--	--	.857	.665	1.098	.228	.945	.729	1.220	.669	1.496	.758	2.901	.239
Black Males	--	--	--	--	1.326	.928	1.863	.112	1.258	.871	1.788	.210	1.038	.328	2.916	.947
Black Females	--	--	--	--	.874	.398	1.710	.714	.860	.385	1.725	.691	.131	.004	1.513	.174
Hispanic Males	--	--	--	--	1.032	.661	1.560	.885	1.206	.763	1.850	.405	1.253	.374	3.668	.697
Hispanic Females	--	--	--	--	1.303	.553	2.733	.510	1.561	.646	3.375	.285	.603	.035	4.984	.683
Interaction Terms																
WF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.886	.754	1.044	.146
BM*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.046	.827	1.343	.715
BF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.511	.900	2.999	.166
HM*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.991	.754	1.323	.951
HF*NCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.293	.752	2.518	.391
N	2,753															

Notes: Estimates with $p < .001$ are bolded in the table. "WF" = White Female; "BM" = Black Male; "BF" = Black Female; "HM" = Hispanic Male; "HF" = Hispanic Female.

Table 5

Differential Prediction of the NVCA Tool, by Race-Ethnicity and Sex Grouping

DV: New Violent Criminal Arrest	Model 1				Model 2				Model 3				Model 4			
	OR	95%CI		*p-value*	OR	95%CI		*p-value*	OR	95%CI		*p-value*	OR	95%CI		*p-value*
County 1																
NVCA Score	**1.624**	**1.554**	**1.697**	**< .001**	--	--	--	--	**1.594**	**1.525**	**1.666**	**< .001**	**1.708**	**1.444**	**2.013**	**< .001**
White Females	--	--	--	--	.563	.366	.837	.006	.604	.392	.901	.017	.284	.104	.712	.010
Black Males	--	--	--	--	**1.625**	**1.346**	**1.978**	**< .001**	1.405	1.161	1.714	.001	1.729	1.136	2.673	.012
Black Females	--	--	--	--	1.175	.947	1.466	.149	1.152	.927	1.440	.208	1.235	.765	2.014	.392
Interaction Terms																
WF*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.422	.980	2.064	.062
BM*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.911	.767	1.085	.291
BF*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.967	.794	1.180	.742
N	19,946															
County 2																
NVCA Score	**1.750**	**1.573**	**1.947**	**< .001**	--	--	--	--	**1.745**	**1.567**	**1.943**	**< .001**	**1.733**	**1.498**	**2.004**	**< .001**
White Females	--	--	--	--	.750	.540	1.028	.080	.874	.625	1.206	.421	.503	.236	1.035	.068
Black Males	--	--	--	--	1.062	.779	1.431	.699	.943	.687	1.279	.709	1.408	.676	2.853	.350
Black Females	--	--	--	--	.829	.450	1.411	.516	.818	.441	1.408	.495	1.012	.261	3.402	.985
Interaction Terms																
WF*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.291	.967	1.728	.083
BM*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.852	.648	1.118	.247
BF*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.913	.546	1.491	.718
N	3,638															
County 3																
NVCA Score	**1.607**	**1.373**	**1.875**	**< .001**	--	--	--	--	**1.595**	**1.361**	**1.864**	**< .001**	**1.690**	**1.380**	**2.064**	**< .001**
White Females	--	--	--	--	.820	.496	1.310	.422	.932	.561	1.497	.778	1.344	.446	3.895	.591
Black Males	--	--	--	--	1.689	.914	2.932	.076	1.616	.869	2.826	.108	3.050	.769	10.881	.097
Hispanic Males	--	--	--	--	1.053	.433	2.181	.900	1.114	.456	2.324	.793	.799	.089	4.800	.822
Hispanic Females	--	--	--	--	1.284	.206	4.322	.734	1.503	.240	5.135	.583	2.606	.053	55.217	.562
Interaction Terms																
WF*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.856	.551	1.294	.473
BM*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.780	.476	1.248	.309
HM*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	1.142	.573	2.270	.700
HF*NVCA Score	--	--	--	--	--	--	--	--	--	--	--	--	.778	.125	3.113	.735
N	2,753															

Notes: Estimates with p < .001 are bolded in the table. “WF” = White Female; “BM” = Black Male; “BF” = Black Female; “HM” = Hispanic Male; “HF” = Hispanic Female.